PART II. NUCLEIC ACIDS

THE STRUCTURE OF THE NUCLEIC ACIDS AND RELATED SUBSTANCES

BY

F. H. C. CRICK

The Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge, England

THERE HAVE BEEN a number of recent reviews covering the structure of deoxyribonucleic acid (DNA),[1,2,3] ribonucleic acid (RNA),[4] and the synthetic RNA-like polyribonucleotides.[5] It therefore seems unnecessary to go over this subject in detail again. In this short paper I shall try to review the reported results in a broader way, in an attempt to emphasize those features that the proposed structures have in common. Such an attempt is perhaps a little premature, but sufficient data are now available to make possible a beginning of this sort.

I shall not discuss in detail the method of obtaining structures from the rather meager information contained in X-ray diagrams, but a few points need emphasis. First, it is essential to know the chemical formula of the material being studied. In most cases what is available is not the *exact* chemical formula, but the general one. More precisely, the formula of the backbone is known, and also the manner in which the purines and pyrimidine bases are attached; their precise sequence remains unknown. Fortunately this is not always as serious a handicap as might be expected since, under X rays, at low resolution, one base looks somewhat like another provided the bases occupy roughly the same position. It would be much more serious, for example, if it were not known whether an α or a β glycosidic link were present. In practice, these difficulties have proved mainly of importance in studying RNA, especially at the period when it was thought that this acid might be rather extensively branched.

Second, the X rays show clearly only the *repeating* part of the structure, and they can be used effectively only to study structures that are spatially regular. Moreover, examination by X ray is a poor way to obtain an answer to the question "how much of the material is in the regular structure?"

Third, information from other sources is useful. Thus, the titration data on DNA have proved most informative, as has the infrared dichroism.[6,7] It would be a great advantage to have more studies on optical rotation to confirm or establish the "hand" of the various helices proposed. At the moment we should like very much to know just how much pairing of bases there is in RNA, and between which bases it occurs.

Fourth, it is important to realize that most (although not all) studies must be made on synthetic materials or on materials extracted from their natural context in the cell. Wilkins' work on various intact cells is a brilliant exception to this realization;[8,9] it

is at present a grave anxiety to us that similar studies on RNA in intact cells have not been feasible.

The Chemical Formula

The information we have is well known,[10] and it can be summarized briefly. The phosphate sugar backbones of DNA, RNA, and the RNA-like polymers are all much alike, the only difference being the substitution of an H for an OH in the 2' position of the sugar in the case of DNA. The glycosidic linkage is believed to be always a β one. The bases in the natural materials are similar; the same four bases commonly occur in both, except that, in DNA, uracil is replaced by thymine (5-methyl uracil).

I think it is more interesting to consider for a moment not what we do find in nature, but what we do not find. It might perhaps have been expected, by analogy with polysaccharides, that we would find mixed nucleic acids containing in the same chain both ribose and deoxyribose residues, either arranged in some repeating alternation, or "at random." The absence of these I believe to be an important fact. To a crystallographer the regularity of the backbone suggests that the structure is likely to be stereochemically regular, and leads him to surmise that perhaps this regularity may be of biological importance. From the genetic angle one notes that the information cannot easily be stored in the backbone alone, as it might be if the ribose-deoxyribose sequence were used as a code; any such information as to the backbone must therefore come from configurational variations, and it seems improbable that these could be sufficiently stable to be acceptable to a geneticist. The regularity of the backbones, then, underlines the irregularity of the base sequence, and it strongly suggests that any genetic information in the nucleic acid is carried by the base sequence, and only by this sequence.

We must briefly note the two rules proposed by Chargaff[11,12] for the over-all base composition of natural nucleic acids. They are

$$\text{DNA} \qquad \frac{A}{T} = \frac{C}{G} = 1$$

$$\text{RNA} \qquad \frac{A + C}{U + G} = 1$$

the letters referring to adenine (A), cytosine (C), thymine (T), guanine (G), and uracil (U), on a molar basis.

I should not wish to be drawn into controversy as to how far these rules are valid, except to say that the first is obeyed for most specimens of DNA, while the second one, which seems to apply in many cases

to material from whole cells, appears to be less well founded; it is probably not correct for certain plant viruses. Nor shall I do more than mention the occurrence of bases other than the standard four, since I have recently discussed elsewhere some aspects of the problems they raise.

The Proposed Structures

We can now proceed to an examination of the proposed structures, the most important features of which are set out in TABLE I.

TABLE 1

Material	Number and disposition of chains	Parameter of screw axis	Distance between successive phosphates
DNA { A	2, antiparallel	2.5₅ A., 33°	5.7A.*
DNA { B	2, antiparallel	3.3₆ A., 36°	6.8 A.*
RNA Poly AU Poly AUGC	unknown; probably 2	(probably similar to poly A)	?
(Poly A + poly U)	2 (parallel?)	3.2 to 3.6 A. } about 36°	probably greater than 7 A.
Poly A	2, parallel	3.8₅ A., about 43°	7.1 A.
Poly I	probably more than one	?	?
Poly C	(some structure, but not known)	?	?
Poly U	amorphous	—	—

The abbreviations used in this table are explained in the text.
*I am indebted to Wilkins¹ for this information.

A few comments are necessary on the nature of the materials. Thus DNA "A" and DNA "B" are the two forms of the DNA structure; the latter occurs at higher humidities and in solution, while the former appears at lower humidities. In most specimens the two are interconvertible. No significant differences have been found in the X-ray patterns of DNA from different biological sources, so it is unnecessary to indicate the type of DNA; in fact, most of the X-ray work has been done with calf thymus DNA.

The term "poly" in the table refers to the synthetic polymers first produced by Grunberg-Manago et al.,[13] who used an enzyme system from bacteria. These polymers have the same backbone as natural RNA, but differ in the sequence of bases. In poly A, for example, every base is adenine. Nobody has yet

succeeded in making a satisfactory poly G, so I have omitted it from the table. Poly AU, when synthesized from an equimolar mixture of adenylic acid and uridylic acid, has approximately equal amounts of adenine and uracil, and Heppel et al.[14] have shown that their sequence is random or very nearly so. Poly A + poly U describes the structure formed by mixing preformed poly A with preformed poly U.

X-ray diagrams of RNA from a variety of sources are all very poor, and they are indistinguishable from those given by poly AU and poly AUGC, so these three materials have been listed together.[4,5] Naturally it is not impossible that improved X-ray pictures might show up significant differences.

X-ray work on all these compounds is being actively pursued, but much remains to be done, both on compounds that exist, such as poly I, and on some not yet made, such as poly CU and poly ACU. Alexander Rich discusses elsewhere in these pages later developments in this field, but I shall mention very briefly poly A (and poly A + poly U) to make my subsequent points intelligible.

The structure of poly A has been studied by Rich, J. D. Watson, D. R. Davies, and myself, and also by G. Morgan and R. S. Bear at the Massachusetts Institute of Technology, Cambridge, Mass. After an extensive trial of one-chain structures both groups have decided in favor of a two-chain structure of the type briefly reported by Watson.[4] In this structure the two helical backbones are parallel rather than antiparallel. The adenines are paired — that is, an adenine on one chain is joined by a pair of hydrogen bonds to its neighbor on the other chain; in addition, the NH_2 of the adenine makes a further hydrogen bond to the phosphate of the opposite chain. To do this the bases must be tilted somewhat, and this increases the distance between the bases in the fiber direction from 3.4 to 3.8 A.

This structure, or simple variants of it, seems very reasonable for the fiber, but it is quite unclear whether it persists into solution. All that is known is that fibers that give good X-ray pictures are not regularly obtained (suggesting that the structure may be a precarious one and that the strong 3.8 A. spacing changes to 3.4 A. as the fiber goes into solution). Special physicochemical studies will therefore be necessary to decide whether poly A is single-stranded or double-stranded in dilute solution. The fact that it has a large hypochromicity, even in very dilute solution, suggests that a regular structure of some sort exists under these conditions.

The most interesting recent work on the subject is undoubtedly that of Rich and Davies[15] on poly A + poly U. Imagine that one has a quantity of poly A

in one test tube and an equimolar quantity of poly U in another. What happens when one mixes them together? One immediately notices a considerable increase in viscosity. Warner[16] had previously shown that there was a striking drop in the total U.V. absorption and that, under electrophoresis, the mixture moved as a single component. All this suggests that a new structure has been formed. Good fibers, which are easily drawn, give good X-ray fiber diagrams. One glance — at least to an experienced eye — suggests that the structure present is similar to that of DNA. Rich and Davies have proposed a DNA-like structure, with base-pairing (of the DNA type) between the adenine and uracil of different chains. The structure is a little fatter than DNA. It is not yet clear whether the two chains are parallel or antiparallel. One will perceive immediately that two important deductions can be drawn from this structure:

(1) Two chains can wind around each other in solution to form a double helix. It is hoped that the physicochemistry of this process will be studied actively in the near future.

(2) An RNA backbone can take up a DNA-like configuration.

I must briefly mention our ideas about RNA and the related poly AU and poly AUGC. It is clear that, for the material as we have it in the test tube, the structure is a rather disordered one. It seems to resemble poly A, except that the strong inner ring of the X-ray pattern may imply that some of the phosphates are at a larger radius. There is also a hint, from the absence of reflections around 3⅓ A. spacing near the meridian, that the structure has a diad (or pseudodiad) parallel to the fiber axis, as poly A has. Thus, an entirely reasonable structure would be one with two parallel chains, with occasional base-pairing, but otherwise rather irregular. Unfortunately the data are much too poor to make this more than an informed guess, and further studies on a variety of polynucleotides would be needed before one could repose much confidence in them. Such a structure would not appear to make sense biologically in any obvious way, and it may well be that the configuration of RNA, as we have it, bears little relation to its configuration inside the cell.

I have not included the DNA nucleoproteins in TABLE 1 because they are somewhat irrelevant, and also because Wilkins, elsewhere in this publication, describes recent developments regarding them.

Possible Generalizations

TABLE 1 shows, first, that thus far *no regular structure has been discovered consisting of a single chain.* Poly U appears amorphous; the hypochromicity is very small and it is probably a rather random coil. Poly I has a structure, but it appears to be about 30 A. across, and thus probably involves more than one chain in the structural unit. About poly C we can say only that some structure is present. In all the other cases it seems probable, with varying degrees of certainty, that the structural unit consists of two chains, although the case of poly A in solution is problematical. Watson, Rich, and I spent much time and effort in the attempt to fit the poly A data to a single-chain model, and this attempt has also been made by Morgan and Bear (personal communication). We now believe that the existence of such a model is unlikely, both on stereochemical grounds and also because its existence does not conform with the available X-ray data.

It seems not unreasonable to reach the tentative conclusion that there is no natural repeating configuration for a *single* nucleic acid chain. Nucleic acid, in fact, seems capable of behaving in a regular manner only in the married state. Whether a single chain of nucleic acid can take up a regular configuration if combined with protein is another matter. This may well occur inside tobacco mosaic virus, for example.

Whereas the existence of a regular structure with one chain seems unlikely, it would be rash to conclude that there will *always* be two chains in the structural unit. There is probably room for a third, or even, under some circumstances, for a fourth. It will be interesting to see whether such structures can be found in nature, or produced from the RNA-like polynucleotides.

Another type of structure that has not been found thus far in these materials is a coiled-coil structure of the general type proposed for collagen and α-keratin. There is some early evidence,[17] however, for a micelle structure for DNA gels. It is tempting to speculate that this is based on the fivefold screw axis, which is a subsymmetry element of the tenfold screw axis of the DNA double-helix in the B form.

One might think that my second point would be that all the structures are helical, but this is not really surprising since there is practically no other way to make a regular arrangement of an asymmetric, repeating chain. My point is, rather, that all the helices are somewhat similar. That is, the parameters of the screw axes are astonishingly alike (see TABLE 1) and, moreover, that they are all right-handed (I note, in passing, that the stable configuration for the α-helix has also been shown recently to be right-handed).[18,19,20]

It is not unreasonable to ascribe this mainly to two causes: the tendency of the bases to stack above one another, and the tendency of the backbone to be

rather fully extended, probably because of the mutual repulsion of the charged phosphate groups. It is too early to do more than mention other possible reasons: the glycosidic bond may have to be approximately radial, for example; or there may be a preferred staggering of the atoms of the phosphate-sugar backbone. In any case it certainly would help us considerably in our attack on such difficult problems as the RNA configuration in small viruses and microsomal particles if we were able to establish the fact that the nucleic acid backbone is likely to vary its configuration only within certain limits.

My next point is that thus far we know of no case in which two backbones lie very close together. Here again it seems unlikely that this would happen unless special arrangements were made to reduce the repulsion between the phosphates in the backbone. It is one of the oldest (unpublished) speculations in this field that this might be done with divalent cations; some studies on materials in which these have replaced sodium might be very valuable. I should not be surprised if a DNA configuration were found in which the two backbones were close together because of the divalent ion, and in which the bases, as a consequence, became unpaired. I shall be interested to learn if Wilkins has any information as to whether the bases are actually paired in nucleohistone. The material recently described by Doty[21] might be an interesting one on which this could be tested.

Finally, a word must be said about base-pairing. Donohue[22] has recently enumerated all the possible ways of pairing the standard bases, assuming that the most likely tautomeric forms exist and that at least two hydrogen bonds are made. It is implied, but not clearly stated, in his paper that there is only one satisfactory way of base-pairing for DNA that will fit the X-ray data if all four bases must occur on both chains. This is the way proposed[23] by Watson and myself and recently refined by Pauling and Corey.[24] However, as the latter have pointed out, different arrangements are possible in other materials. In particular the pairing proposed for poly A is that actually found by Broomhead[25] in crystals of adenine hydrochloride. Recently it has been suggested (Bernhard, unpublished, and Geiduschek, unpublished) that mutually induced ionization leading to zwitterion formation may play an important part in the pairing of bases, but this idea must at the moment be regarded as speculative. Clearly, it will be necessary to carry out some fundamental work on the pairing of bases, either by studying it in mixed crystals of the monomers (if such crystals can be found) or between nucleotides or polynucleotides in solution. In passing it should be noted that the pairing of bases proposed by Donohue and Stent[26] in their "mating helix" for RNA involves a pairing not included in Donohue's paper, since it requires a tautomeric shift of one of the hydrogen atoms of the guanines. It is also very unclear whether the bases can be turned over, through 180° about the glycosidic bond, from their most likely positions (those found in DNA). The structure of the nucleic acids is of such fundamental importance that more work on the basic physical chemistry of its components is certain to return valuable dividends.

REFERENCES

1. WILKINS, M. H. F. 1956. Biochem. Soc. Symposia. Cambridge, Eng. In press.

2. LANGRIDGE, R., W. E. SEEDS, H. R. WILSON, C. W. HOOPER, M. H. F. WILKINS & L. D. HAMILTON. 1957. In press.

3. CRICK, F. H. C. 1957. In The Chemical Basis of Heredity. W. D. McElroy and B. Glass, Eds. Johns Hopkins Press. Baltimore, Md.

4. WATSON, J. D. 1957. In The Chemical Basis of Heredity. W. D. McElroy and B. Glass, Eds. Johns Hopkins Press. Baltimore, Md.

5. RICH, A. 1957. In The Chemical Basis of Heredity. W. D. McElroy and B. Glass, Eds. Johns Hopkins Press. Baltimore, Md.

6. SUTHERLAND, G. B. B. M. & M. TSUBOI. 1957. Proc. Roy. Soc. London. In press.

7. JORDAN, D. O. 1955. In The Nucleic Acids. 1:447. E. Chargaff and J. N. Davidson, Eds. Academic Press. New York, N.Y.

8. WILKINS, M. H. F. & J. T. RANDALL. 1953. Biochim. et Biophys. Acta. 10:192.

9. WILKINS, M. H. F., A. R. STOKES & H. R. WILSON. 1953. Nature. 171:738.

10. THE NUCLEIC ACIDS. 1955. 1. E. Chargaff and J. N. Davidson, Eds. Academic Press. New York, N. Y.

11. CHARGAFF, E. 1955. *In* The Nucleic Acids. 1:307. E. Chargaff and J. N. Davidson, Eds. Academic Press, New York, N. Y.

12. ELSON, D. & E. CHARGAFF. 1955. Biochim. et Biophys. Acta. 17:367.

13. GRUNBERG-MANAGO, M., P. J. ORTIZ & S. OCHOA. 1956. Biochim. et Biophys. Acta. 20:269.

14. HEPPEL, L. A., P. J. ORTIZ & S. OCHOA. In preparation.

15. RICH, A. & D. R. DAVIES. 1956. J. Am. Chem. Soc. 78:3548.

16. WARNER, R. C. 1956. Federation Proc. 15:379.

17. RILEY, D. P. & G. OSTER. 1951. Biochim. et Biophys. Acta. 7:526.

18. YANG, J. T. & P. DOTY. 1957. J. Am. Chem. Soc. In press.

19. ELLIOTT, A. & E. R. MALCOLM. 1956. Nature. 178:912.

20. ELLIOTT, A., W. E. HANBY & E. R. MALCOLM. 1956. Nature. 178:1170.

21. DOTY, P. & G. ZUBAY. 1957. In press.

22. DONOHUE, J. 1956. Proc. Natl. Acad. Sci. 42:60.

23. CRICK, F. H. C. & J. D. WATSON. 1954. Proc. Roy. Soc. London. A223:80.

24. PAULING, L. & R. B. COREY. 1956. Arch. Biochem. Biophys. 65:164.

25. BROOMHEAD, J. M. 1948. Acta Cryst. 1:324.

26. DONOHUE, J. & G. S. STENT. 1956. Proc. Natl. Acad. Sci. 42:734.